

IN THE CLAIMS

This listing of the claim will replace all prior versions and listings of claim in the present application.

Listing of Claims

Claims 1 and 2 (canceled).

3. (currently amended) A similar document search method, implemented in a document search and retrieval system including a memory for storing a document database having documents and information concerning the documents and a processor for processing data of the document database, of searching a document similar to a seeds document from said a document database, the method, comprising the steps executed by the processor, of:

~~a first extracting step of extracting at least one characteristic word from the seeds document including desired retrieval contents;~~

~~a second extracting step of extracting as characteristic words of the seeds document, if the at least one characteristic word extracted by the first extracting step is one of a compound characteristic phrase constructed by a plurality of constituent characteristic words included in the compound characteristic phrase, second extracting of characteristic words included in said compound characteristic phrase and each of said plurality of constituent characteristic words of said compound, and the constituent characteristic words included in the compound characteristic phrase;~~

~~a step of calculating, according to the characteristic words extracted by the second extracting step, similarity between the seeds document and the document stored on the document database, by using the characteristic~~

words including the compound characteristic phrase and the constituent characteristic words by which the compound characteristic phrase is constructed; and

~~a step of outputting a retrieval result as a result of the similarity calculated by the similarity calculating step,~~

wherein the similarity calculating step includes the steps of:

~~a step of calculating a weighting coefficient corresponding to a distance between said constituent characteristic words on the document stored on the document database, said distance being calculated based on term appearance positions of said constituent characteristic words in a phrase including said characteristic words extracted in said second extracting step on the document stored on the document database; calculated by term appearance position on the seeds document between a constituent characteristic word and another constituent characteristic word which are extracted from one compound characteristic phrase, and~~
~~a step of calculating similarity involving by multiplying the weighting coefficient.~~

Claims 4 and 5 (canceled).

6. (currently amended) A similar document search system for searching a document similar to a seeds document from a document database, the system, comprising:

a memory for storing the document database having documents and information concerning the documents; and

a processor for processing data of the document database,

wherein said processor comprises:

a document analyzer processor for extracting at least one characteristic word from the seeds document including desired retrieval contents;

a characteristic word extractor processor for ~~extracting as characteristic words of the seeds document~~, if the at least one characteristic word extracted by the document analyzer processor is one of a compound characteristic phrase constructed by a plurality of constituent characteristic words included in the compound characteristic phrase, extracting characteristic words included in said compound characteristic phrase and each of said plurality of constituent characteristic words of said compound, ~~and the constituent characteristic words included in the compound~~ characteristic phrase;

a seeds document similarity calculator processor for calculating, according to the characteristic words extracted by the characteristic word extractor processor, similarity between the seeds document and the document stored on the document database, by using the characteristic words including the compound characteristic phrase and the constituent characteristic words by which the compound characteristic phrase is constructed; and

a retrieval result output processor for outputting a retrieval result as a result of the similarity calculated by the seeds document similarity calculator processor,

wherein the seeds document similarity calculator processor includes:

a weighting coefficient calculator processor for calculating a weighting coefficient corresponding to a distance between said constituent characteristic words on the document stored on the document database, said distance being calculated based on term appearance positions of said constituent characteristic words in a phrase including said characteristic words extracted in said second extracting step on the document stored on the document database~~calculated by term appearance position on the seeds document between a constituent characteristic word and another constituent characteristic word which are extracted from one compound characteristic phrase, and~~

a calculator processor for calculating similarity involving by multiplying the weighting coefficient.

Claims 7 and 8 (canceled).

9. (currently amended) A program, stored on a storage medium, executable by a document search and retrieval system including a memory for storing a document database having documents and information concerning the documents and a processor for processing data of the document database ~~product for making a computer operate as a similar document search system for searching a document similar to a seeds document from said a document database, the program, when executed cause the document search and retrieval system, to perform the steps comprising of:~~

~~first a document analyzer processor program for extracting at least one~~
characteristic word from the seeds document including desired retrieval
contents;

~~a characteristic word extractor processor program for extracting as~~
~~characteristic words of the seeds document, if the at least one characteristic~~
word extracted by the first extracting step ~~document analyzer processor~~
~~program is one of a compound characteristic phrase constructed by a plurality~~
~~of constituent characteristic words included in the compound characteristic~~
~~phrase, second extracting of characteristic words included in said compound~~
~~characteristic phrase and each of said plurality of constituent characteristic~~
~~words of said compound, and the constituent characteristic words included in~~
~~the compound characteristic phrase;~~

~~a seeds document similarity calculator processor program for~~
calculating, according to the characteristic words extracted by the second
extracting step ~~characteristic word extractor processor program, similarity~~
between the seeds document and the document stored on the document
database, by using the characteristic words including the compound
characteristic phrase and the constituent characteristic words by which the
compound characteristic phrase is constructed; and

~~a retrieval result output processor program for outputting a retrieval~~
result as a result of the similarity calculated by the calculating step ~~seeds~~
~~document similarity calculator processor program,~~

wherein the calculating step ~~seeds document similarity calculator~~
~~processor program includes the steps of:~~

~~a weighting coefficient calculator processor program for calculating a~~
weighting coefficient corresponding to a distance calculated by term
appearance position on the seeds document between a constituent
characteristic word and another constituent characteristic word which are
extracted from one compound characteristic phrase, and

~~a calculator processor program for calculating similarity involving by~~
~~multiplying the weighting coefficient.~~